






# Should low-quality evidence dominate high-level evidence? A systematic review and meta-analysis of systematic reviews of musculoskeletal physical therapy interventions

Sean Riley <sup>a</sup>, Brian T. Swanson<sup>b</sup>, Steven F. Sawyer<sup>c</sup>, Jean-Michel Brismée <sup>c</sup> and Geoffrey Staysniak <sup>a</sup>

<sup>a</sup>Physical Therapy Program, Sacred Heart University, Fairfield, CT, USA; <sup>b</sup>Department of Rehabilitation Sciences, University of Hartford, West Hartford, CT, USA; <sup>c</sup>Department of Rehabilitation Sciences and Center for Rehabilitation Research, Texas Tech University Health Sciences Center, Lubbock, TX, USA

## ABSTRACT

**Objectives:** To determine the: 1) quality of articles cited in systematic reviews (SRs); 2) methodological quality of the SRs; and 3) impact of quality on level 1A evidence.

**Methods:** SRs related to musculoskeletal physical therapy interventions were identified. The methodological quality of the SRs and articles cited by the SRs were assessed by two blinded reviewers. Data analysis was performed by a third blinded researcher. Additional comparisons were made based on the Journal Impact Factor, spin, financial bias, and conflict of interest.

**Results:** Twenty-four SRs were identified; 21/24 SRs had 'critically low' quality on the AMSTAR 2. Thirty-four percent of included studies were 'low quality,' and 58% of SRs included studies that had unreported external validity. One-half of the SRs represented 'spin,' and one-third of the SRs generated conclusions based on low-quality clinical trials.

**Discussion:** The 'critically low' SRs methodological quality was exacerbated by low-quality research inclusion. Most SRs failed to follow best practices, including prospective registration and integration of professional librarians in the search process. Based on the high proportion of SRs that include low-quality trials and overall low methodological quality, further discussion regarding practice recommendations on level vs. quality of evidence is warranted.

**Level of Evidence:** 1a

## KEYWORDS

Journal impact factor; physical therapists; publications; randomized controlled trials as topic; bias

## Introduction

Systematic reviews (SRs) of randomized clinical trials (RCTs), with and without meta-analyses, are considered to be the highest level of evidence in research (Level 1A) [1]. Level 1A evidence relates to Grade A practice recommendation, which is a strong recommendation that clinicians should follow unless they have a 'clear and compelling' reason for using alternative approaches [1]. An SR in 2019 found that the median percentage of physical therapists' use of interventions supported by evidence-based guidelines was 54% to 63% [2]. This study found that treatments that were not recommended were provided 20–37% of the time, and treatments with no recommendations were provided 45%–81% of the time [2]. Although Cochrane reviews report conclusive results in other health-care fields 45%–80% of the time [3–8], they provide conclusive results in only 4.7% of reviews when examining physical therapists' interventions [9]. This result is primarily related to the proliferation of poorly designed clinical trials [9].

It has been recognized that the perspectives, allegiances, and biases of the team that conducts an SR can have a significant impact on the SR outcome [10]. The question then becomes, if the team that performs the SR significantly impacts the results, to what extent

does this occur when assessing the quality of RCTs included in the SR, and what impact does 'spin' have on the SR outcome? 'Spin' has been defined as 'specific reporting strategies used by authors to convince readers that the beneficial effect of the treatment of interest is greater than shown by the results' [11]. Although there is significant scrutiny placed on primary evidence and RCTs regarding affiliation, secondary syntheses such as SRs with or without meta-analyses do not undergo the same level of scrutiny before publication [10].

Considering the prevalence of poorly designed clinical trials involving physical therapy interventions, how should these studies be considered during the systematic review (SR) process, particularly in the presence of high-quality clinical trials? Combining any volume of low-quality studies with high-quality studies during the SR process is sure to influence results, alter effect sizes, and, at worst, may allow poorly conducted and biased research to dominate the SR and meta-analytical process. When this occurs, such SRs allow low-quality research to masquerade as high-quality research, which wastes valuable resources and substantially hinders progress within the physical therapy profession. It, therefore, becomes essential to identify variables that may methodologically introduce bias and influence the SRs quality.

The objectives of this methodological SR were to: (1) identify SRs of RCTs published in the English language related to musculoskeletal physical therapy (MSKPT) interventions; (2) use the PEDro (Physiotherapy Evidence Database) website to determine the quality of the RCTs used in SRs; (3) determine the impact factor of the journals in which the SRs were published; (4) assess the quality of the identified SRs using A Measurement Tool to Assess Systematic Reviews, version 2 (AMSTAR 2) (5) evaluate whether SR authors employed 'spin' by over estimating the outcome and reporting a conclusion inconsistent with the results; (6) determine if the SR used a professional librarian to create the literature search strategy; (7) determine if the journal that published the SR was financially biased; (8) attempt to determine if the team conducting the SR had any unreported conflicts of interest; (9) assess differences between PEDro scores reported in the SRs and certified PEDro scores (<https://www.pedro.org.au/>); and (10) determine if there were relationships between the metrics of quality of the articles reviewed by the SRs and the quality metrics of the SRs.

## Methods

### *Protocol and registration*

The aim of this SR was to create an unbiased sample of SRs, representing commonly addressed musculoskeletal symptoms in physical therapy across a broad range of journals widely used by physical therapists. Since this is a systematic methodology review without direct patient outcome or clinical relevance, there was no registration in PROSPERO. This methodology has been previously published in methodological SRs [12,13].

### *Definition of systematic reviews and meta-analyses*

The guidelines provided by the Cochrane Collaboration and the PRISMA Statement were used to outline this methodological SR and to define what we classified as SRs and meta-analyses [14,15].

### *Sample size determination and eligibility criteria*

The sample for this study was created for a methodological review to determine which factors contained within an SR had the most significant impact on the quality of the SR. Searches were limited to SRs of RCTs published in English related to musculoskeletal physical therapy (MSKPT) interventions from 1 January 2016 to 12 December 2019 (the last search date). This date range was selected in order to obtain studies that had the opportunity to use the most recent CONSORT statement and to capture a time frame during which journals

endorsed its use [16]. We also wanted to capture a time frame that mainly included the update to the AMSTAR 2 recommendations that were published in 2017 [17]. Additionally, we tried to critically appraise SRs that were high in methodological quality and determine the impact of low quality and top-quality RCTs on this quality. It has been suggested that SRs of RCTs are of higher methodological quality secondary to the limited methodology and nature of the disciplined research design of the studies included in this group of SRs [18]. Finally, MSKPT interventions are a specific interest related to the authors' clinical practice. To be eligible for inclusion in this study, an SR had to use the PEDro score to assess the quality of the included RCTs. Studies were excluded if they used multiple tools to assess the RCTs included in the SR to create a homogeneous sample and prevent the addition of potential confounding variables into the data analysis and interpretation.

### *Information sources*

The methods outlined by Furlan [19] and Lefebvre [20] were used to guide the development of search strategies. Controlled vocabulary and free text terms related to physical therapy musculoskeletal interventions were used to search the following databases and registries for SRs: PubMed, which includes MEDLINE (1 January 2016–12 December 2019), and CINAHL.

### *Search*

The search strategies are presented in Appendix 1. These database searches were supplemented by citation searching using bibliographies of reviews and research articles that met the inclusion and exclusion criteria.

The search strategy was developed by a professional librarian experienced in systematic literature reviews, and peer-reviewed by another librarian using the PRESS standard [21]. The use of a professional librarian to generate the search strategy is an essential methodological quality consideration when performing an SR [22–25].

### *Study selection*

Search results were imported to EndNote and screened by title and then imported to Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia; [www.covidence.org](http://www.covidence.org)) for further review. Abstracts were screened by two of the authors (SPR, BTS), and in the event of disagreement, a consensus was reached through discussion of the conflicting abstract. A full-text review was then undertaken in duplicate.

### **Assessment for bias in PEDro scale scoring**

The certified PEDro scale scores for the RCTs reviewed by the SRs were attained from the PEDro website (<https://search.pedro.org.au/search>). Only RCTs listed on the PEDro website were used for statistical analysis of PEDro scores. Previous assessments have determined that the PEDro scale is a reliable [26,27] and valid [28] assessment of RCT quality. The PEDro score consists of 11 criteria, with the first criterion representing the external validity of the study [29]. The second through eleventh criteria is therefore reported as the PEDro Score, which can range from 0 to 10 [29]. The kappa values for individual scale items have been shown to range from 0.50 to 0.79, with an intraclass correlation coefficient value of 0.68, 95% CI = 0.57, 0.76 for the total score [26]. The PEDro scale has also been shown to display a strong correlation with the Cochrane Risk of Bias (CRoB) tool ( $r = 0.83$ ; 95% confidence interval [CI] 0.76 to 0.88) [30] and therefore may be considered a study's risk of bias assessment [31].

### **Assessment of systematic review quality**

The AMSTAR 2 tool was used to assess the methodological quality of systematic reviews ([https://amstar.ca/Amstar\\_Checklist.php](https://amstar.ca/Amstar_Checklist.php)). The AMSTAR 2 contains 16 questions. Of the 16 questions, seven are considered critical to the validity of the review and its conclusions [17]. These critical domains consist of: 'Item 2 – Protocol registered before commencement of the review; Item 4 – Adequacy of the literature search; Item 7 – Justification for excluding individual studies; Item 9 – Risk of bias from individual studies being included in the review; Item 11 – Appropriateness of meta-analytical methods; Item 13 – Consideration of risk of bias when interpreting the results of the review; and Item 15 – Assessment of presence and likely impact of publication bias [17].' The rating of the overall confidence in the review is rated as high, moderate, low, and critically low, depending on the strength and weaknesses of the available studies [17].

Two independent, blinded raters independently assessed the identified systematic reviews. These assessments were then sent to a third reviewer to determine disagreements in the ratings. These disagreements were then sent back to the initial reviewers to decide whether they could come to a consensus. If they could not come to an agreement, the third reviewer served as a tie-breaker [32].

### **Assessment of the influence of the team conducting a systematic review**

Every effort was made to determine if the team conducting the SR had a financial or personal conflict of interest in the outcome of the systematic review. A google search was performed for each author by

name. The results were assessed for potential conflicts of interest (commercial products, for-profit educational seminars, etc.). These results were compared to the conflict of interest disclosures contained in the SR. Conflict of interest and financial interest were assessed as yes, no, or not disclosed.

### **Assessment of spin**

SRs were evaluated for Spin based on the seven criteria established by Nascimento et al. [18]. These criteria were: '1) the conclusion formulates recommendations for clinical practice not supported by the findings; 2) the title claims or suggests a beneficial effect of the experimental intervention not supported by the findings; 3) selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention; 4) the conclusion claims safety based on non-statistically significant results with a wide confidence interval; 5) the conclusion claims a beneficial effect of the experimental treatment despite a high risk of bias in primary studies; 6) selective reporting of, or overemphasis on, harm outcomes or analysis favoring the safety of the experimental intervention; and 7) the conclusion extrapolates the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions) [18].' Consistent with the methodology of Nascimento et al., each item was judged as having Spin if it was present or information was omitted or not reported. If the SR was determined to have fulfilled the category 'Claims benefit despite a high risk of bias in primary studies,' we only considered this to be positive in the presence of a second positive category. This was done to give the authors the benefit of the doubt given that as high as 90% of published SRs have been rated as low or critically low [33].

These variables were assessed by the two raters in the same manner as the AMSTAR 2 scores.

### **Assessment of the influence of financially biased journals**

Financially Biased Journals (predatory) were identified based on the criterion described by the publishing industry initiative at [thinkchecksubmit.org](http://thinkchecksubmit.org). This included journals that (i) required payment for publication and (ii) were not members of any of the following organizations: the Committee on Publication Ethics (COPE), Directory of Open Access Journals (DOAJ), Open Access Scholarly Publishers' Association (OASPA), International Network for the Availability of Scientific Publications (INASP), Journals Online platforms (for journals published in Bangladesh, Nepal, Sri Lanka, Central America, and Mongolia), or African Journals Online (AJOL, for African journals), and (iii) did not have an impact factor [34].

## Assessment of the influence of journal impact factor

Journal Impact factors (JIF) for the year of publication were attained through the 2019 Journal Citation Reports, Clarivate Analytics, 2019. (Available at <https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/>)

## Synthesis of results

Data were collated in a Microsoft Excel spreadsheet. Descriptive statistics and inferential analyses were performed with MedCalc Statistical Software version 19.0.7 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>) and GraphPad InStat version 3.10 for Windows (GraphPad Software, San Diego California USA, [www.graphpad.com](http://www.graphpad.com)). For multiple comparisons of correlation analysis, alpha adjustments were performed using the step-wise approach of Hochberg [35].

## Results

### Study selection

A PRISMA diagram is included in Figure 1. SRs identified through database searching included 2745, and 1

additional SR that was identified through a manual search of references. There were 2631 SRs excluded after screening the titles. One-hundred and 15 abstracts were screened, and 41 were excluded. Seventy-four full-text articles were assessed for eligibility, and 50 were excluded. The reasons for these exclusions are included in Figure 1. Five SRs were excluded that used a single tool for critical appraisal that was not the PEDro score. Three were excluded that modified a single RCT critical appraisal instrument. There were also 10 different combinations of individual instruments used to appraise RCTs in the identified SRs critically. This left 24 SRs included in the narrative synthesis that utilized the PEDro as the only instrument for critical appraisal of the RCTs in the identified SRs.

### Study characteristics

The literature citations of the 24 SRs that met eligibility for inclusion in this study are listed in Table 1.

### Quality of articles cited in SRs

A summary of attributes of the studies that were reviewed in the 24 SRs is presented in Table 2. The number of articles reviewed by the SRs varied widely, from 3 to 45 (mean  $\pm$  SD: 13.7  $\pm$  10.2), and a median of

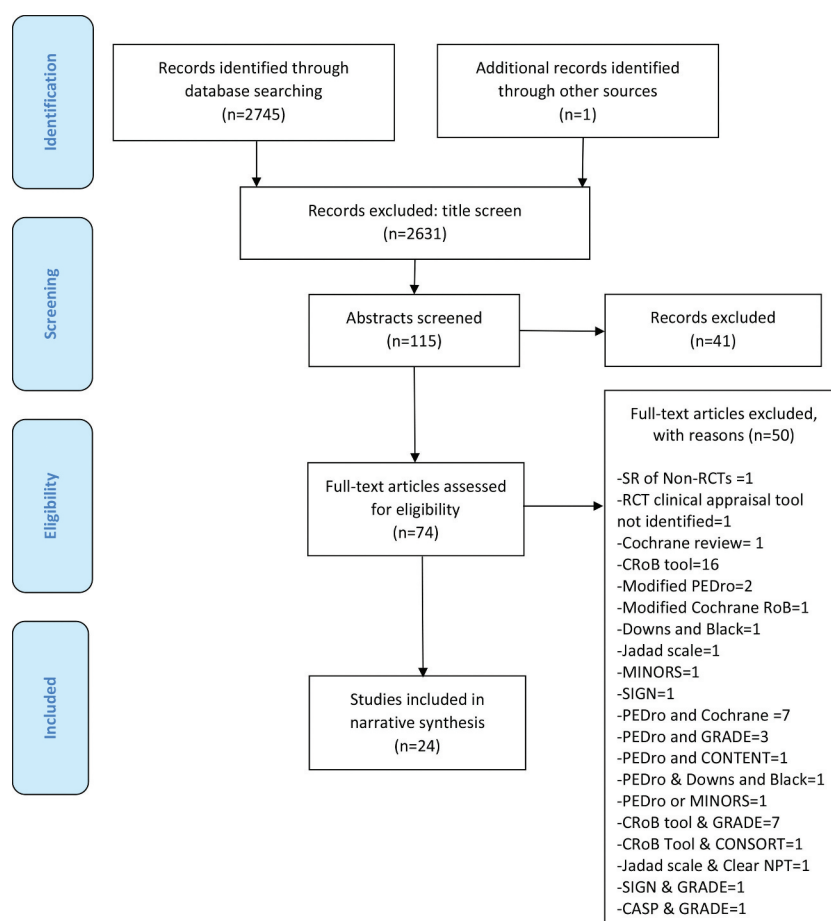


Figure 1. PRISMA diagram.

**Table 1.** Listing of systematic review articles that met criteria for inclusion in this study. Listing of studies is based on the alphabetic order of the first author.

SR Identifier	Authors	Title	Journal	Year	Volume	Pages
1	Amiri et al.	The effect of different exercise programs on size and function of deep cervical flexor muscles in patients with chronic nonspecific neck pain: A systematic review of randomized controlled trials	<i>Am J Phys Med Rehabil</i>	2017	96	582–588
2	Ballester-Perez et al.	Effectiveness of nerve gliding exercises on carpal tunnel syndrome: A systematic review	<i>J Manipulative Physiol Ther</i>	2017	40	50–59
3	Bury et al.	Effectiveness of scapula-focused approaches in patients with rotator cuff related shoulder pain: A systematic review and meta-analysis	<i>Man Ther</i>	2016	25	35–42
4	Byrnes et al.	Is Pilates an effective rehabilitation tool? A systematic review	<i>J Bodyw Mov Ther</i>	2018	22	192–202
5	Cayco et al.	Hold-relax and contract-relax stretching for hamstrings flexibility: A systematic review with meta-analysis	<i>Phys Ther Sport</i>	2019	35	42–55
6	Ceballos-Laita et al.	Effects of non-pharmacological conservative treatment on pain, range of motion and physical function in patients with mild to moderate hip osteoarthritis. A systematic review	<i>Complement Ther Med</i>	2019	42	214–222
7	Diong et al.	Structured exercise improves mobility after hip fracture: A meta-analysis with meta-regression	<i>Br J Sports Med</i>	2016	50	346–355
8	Eisele et al.	Behavior change techniques applied in interventions to enhance physical activity adherence in patients with chronic musculoskeletal conditions: A systematic review and meta-analysis	<i>Patient Educ Couns</i>	2019	102	25–36
9	Fredin and Loras	Manual therapy, exercise therapy or combined treatment in the management of adult neck pain – A systematic review and meta-analysis	<i>Musculoskelet Sci Pract</i>	2017	31	62–71
10	Girard and Girard	The effects of qigong on neck pain: A systematic review	<i>Complement Ther Clin Pract</i>	2019	34	23–29
11	Griffin et al.	General exercise does not improve long-term pain and disability in individuals with whiplash-associated disorders: A systematic review	<i>J Orthop Sports Phys Ther</i>	2017	47	472–480
12	Honda et al.	Effects of physical-agent pain relief modalities for fibromyalgia patients: A systematic review and meta-analysis of randomized controlled trials	<i>Pain Res Manag</i>	2018	Article ID 2,930,632, 9 pages	
13	Logan et al.	Systematic review of the effect of taping techniques on patellofemoral pain syndrome	<i>Sports Health</i>	2017	9	456–461
14	Louw et al.	The efficacy of pain neuroscience education on musculoskeletal pain: A systematic review of the literature	<i>Physiother Theory Pract</i>	2016	32	332–355
15	Mattos et al.	Effects of aquatic exercise on muscle strength and functional performance of individuals with osteoarthritis: a systematic review	<i>Rev Bras Reumatol Engl Ed</i>	2016	56	530–542
16	Meyer et al.	Secondary prevention of chronic musculoskeletal pain: A systematic review of clinical trials	<i>Ann Phys Rehabil Med</i>	2018	61	323–338
17	Mine et al.	Effectiveness of stretching on posterior shoulder tightness and glenohumeral internal-rotation deficit: A systematic review of randomized controlled trials	<i>J Sport Rehabil</i>	2017	26	294–305
18	Nascimento et al.	Hip and knee strengthening is more effective than knee strengthening alone for reducing pain and improving activity in individuals with patellofemoral pain: A systematic review with meta-analysis	<i>J Orthop Sports Phys Ther</i>	2018	48	19–31
19	O'Keefe et al.	Are group-based and individual physiotherapy exercise programmes equally effective for musculoskeletal conditions? A systematic review and meta-analysis	<i>Br J Sports Med</i>	2017	51	126–132
20	Ortega-Castillo and Medina-Porques	Effectiveness of the eccentric exercise therapy in physically active adults with symptomatic shoulder impingement or lateral epicondylar tendinopathy: A systematic review	<i>J Sci Med Sport</i>	2016	19	438–453
21	Reijnveld et al.	Clinical outcomes of a scapular-focused treatment in patients with subacromial pain syndrome: A systematic review	<i>Br J Sports Med</i>	2017	51	436–441
22	Wewege et al.	Aerobic vs. resistance exercise for chronic nonspecific low back pain: A systematic review and meta-analysis	<i>J Back Musculoskelet Rehabil</i>	2018	31	889–889
23	Young et al.	The influence of exercise dosing on outcomes in patients with knee disorders: A systematic review	<i>J Orthop Sports Phys Ther</i>	2018	48	146–161
24	Young et al.	The influence of dosing on effect size of exercise therapy for musculoskeletal foot and ankle disorders: A systematic review	<i>Braz J Phys Ther</i>	2018	22	20–32



**Table 2.** Quality of articles in systematic reviews, based on criteria listed in table headers. Refer to the text for additional explanation.

SR Identifier (Table 1)	Median JIF (2019) of Journals that Published Articles Reviewed by SR	Number of Articles Reviewed by SR	% of Articles Reviewed by SR Published in Journals without JIF	SR Report of PEDro scores (mean values)	PEDro: Mean PEDro score (max- imum = 10) of Articles Reviewed by SR (from PEDro website)	PEDro: % of Articles Reviewed by SR that were Low Quality (PEDro score < 6)	PEDro: Articles Reviewed by SR that did not meet PEDro Criterion 1 (Eligibility criteria, External Validity)?	SR Included Pilot Studies?	SR Included Non-Peer Reviewed Articles?
1	0.814	9	44%	5.89	6.00	56%	No	No	No
2	1.908	13	8%	6.00	5.44	62%	No	No	Yes
3	0.407	4	50%	5.75	5.50	75%	Yes	No	No
4	1.197	23	43%	5.48	5.80	52%	Yes	No	No
5	2.000	39	28%	4.62	3.70	36%	Yes	Yes	No
6	2.697	12	8%	7.17	7.10	17%	No	No	Yes
7	2.738	19	26%	6.37	6.40	21%	Yes	No	No
8	3.188	22	0%	6.82	6.82	18%	No	No	Yes
9	2.622	7	14%	6.00	6.30	57%	No	No	No
10	2.903	5	20%	6.40	5.17	40%	Yes	No	No
11	4.466	3	0%	8.00	8.00	0%	No	No	No
12	2.200	11	0%	6.27	5.70	45%	Yes	No	No
13	3.058	5	0%	7.43	7.20	20%	No	No	No
14	2.697	13	23%	8.23	6.60	8%	Yes	No	No
15	2.330	12	0%	8.00	6.80	8%	Yes	No	Yes
16	3.043	13	0%	5.31	5.50	54%	Yes	Yes	Yes
17	2.032	10	40%	6.10	5.00	40%	Yes	No	No
18	2.720	14	7%	5.80	5.92	36%	Yes	No	Yes
19	2.802	14	7%	6.57	6.43	21%	No	Yes	No
20	2.802	12	17%	6.00	5.83	25%	Yes	Yes	No
21	1.284	4	25%	6.75	5.50	50%	Yes	Yes	Yes
22	2.495	6	0%	5.67	5.67	50%	No	Yes	No
23	2.738	45	7%	6.04	6.00	36%	Yes	Yes	No
24	3.361	14	7%	6.86	6.86	14%	No	No	No
<b>Mean</b>	<b>2.437</b>	<b>13.7</b>	<b>15.6%</b>	<b>6.40</b>	<b>6.05</b>	<b>35.0%</b>	–	–	–
<b>Median</b>	<b>2.697</b>	<b>12.0</b>	<b>8.0%</b>	<b>6.19</b>	<b>5.96</b>	<b>35.8%</b>	–	–	–
<b>S.D.</b>	<b>0.873</b>	<b>10.2</b>	<b>16.1%</b>	<b>0.89</b>	<b>0.88</b>	<b>19.7%</b>	–	–	–
<b>Range</b>	<b>0.407 to 4.466</b>	<b>3 to 45</b>	<b>0% to 50%</b>	<b>4.62 to 8.23</b>	<b>3.70 to 8.0</b>	<b>0% to 75%</b>	–	–	–
<b>Frequency Counts</b>	–	–	–	–	–	–	<b>14 Yes 10 No</b>	<b>7 Yes 17 No</b>	<b>7 Yes 17 No</b>

JIF = Journal Impact Factor; PEDro = Physiotherapy Evidence Database (1 to 10 scale; higher value represents better study design).

AMSTAR 2 = A Measurement Tool to Assess systematic Reviews, version 2 (16 criteria; high value represents better review design).

SR = Systematic Review.

12.0. There were no significant correlations between the number of articles and quality metrics (Table 4).

The Journal Impact Factor (JIF) for the journals from which the reviewed articles were published varied in a non-normal fashion in many SRs (mostly low JIF values mixed with a few high JIF values), so median values are reported in Table 2. The median JIF values ranged from 0.41 to 4.47 ( $2.44 \pm 0.86$ ). Many of the SRs (17/24) included articles published in journals with no JIF value, with seven SRs having 25% or more of such articles.

PEDro scores range from 0 to 10 based on 10 internal validity metrics (PEDro criteria #2 to #11), with articles scored at less than six being considered 'low quality.' The mean of PEDro website scores (i.e., the number of metrics met) of the reviewed articles within the SRs ranged from 3.70 to 8.00 ( $6.05 \pm 0.86$ ). All but one SR included 'low quality' articles in their review; the sole exception only reviewed three articles. For 14 (58%) of the SRs, one-third or more of the reviewed articles were low quality. Collectively for the 24 SRs, 111 out of the 329 (34%) reviewed articles were low quality. More than one-half of the SRs (14/24) included articles that did not meet the PEDro metric for external validity (criterion #1).

Furthermore, seven of the SRs (29.1%) included articles that were pilot studies, and seven SRs included articles that were not peer-reviewed.

### Quality of the SRs

A summary of the attributes of the quality of the 24 SRs is presented in Table 3. Using the 16 domains of AMSTAR 2 as a measure of the quality of the SRs, for the 24 SRs, there was a range of 7 to 13 of the 16 domains being met or partially met ( $10.4 \pm 1.7$ ). Of the 24 SRs, 3 were 'low' confidence, and 21 were considered to be 'critically low'; of the latter, 16 SRs had three or four critical validity flaws.

The criteria for having 'Spin' were met in one-half (12/24) of the SRs, and one-third (8/24) of the SRs made conclusions about the literature that were based in part on low-quality articles (using PEDro definition described above and employed in Table 2). Our modification of the criteria established by Nascimento et al. [18] where we only considered Spin to be present if the category 'Claims benefit despite the high risk of bias in primary studies' was present with a second positive category resulted in 2 studies that would otherwise be

**Table 3.** Quality of systematic review, based on criteria listed in table headers. Refer to the text for additional explanation.

SR Identifier (Table 1)	JIF (2019) of Journal that Published SR	Number of Reviewed Articles by SR	AMSTAR 2: Number of 16 criteria met or partially met by SR	AMSTAR 2: Number of 7 Critical Validity Domain criteria met or partially met by SR	AMSTAR 2: Rating of SR (based on Critical Validity Domain criteria)	SR had "Spin" (at least 2 of 7 criteria met)?	SR Conclusions based in part on Reviewed Low Quality Articles?	SR Included Conflict of Interest Statement?	SR Used Librarian for Literature Search?	Indication of Financial Bias by SR Authors?
1	1.908	9	9	4	Critically Low	No	No	Yes	No	No
2	1.274	13	8	4	Critically Low	No	No	Yes	No	No
3	2.622	4	12	6	Low	Yes	No	Yes	No	No
4	1.120	23	9	3	Critically Low	Yes	Yes	Yes	No	No
5	2.000	39	12	5	Critically Low	Yes	Yes	Yes	No	No
6	1.979	12	11	4	Critically Low	Yes	No	Yes	No	No
7	11.645	19	12	5	Critically Low	No	No	Yes	No	No
8	2.821	22	13	6	Low	Yes	No	No	No	No
9	1.725	7	11	5	Critically Low	Yes	Yes	Yes	No	No
10	1.587	5	7	3	Critically Low	Yes	No	Yes	No	No
11	3.058	3	11	4	Critically Low	Yes	No	Yes	No	No
12	1.701	11	10	3	Critically Low	Yes	Yes	Yes	No	No
13	2.649	5	10	4	Critically Low	No	No	Yes	No	No
14	1.158	13	8	3	Critically Low	Yes	No	No	No	No
15	1.163	12	10	3	Critically Low	Yes	No	Yes	No	No
16	4.196	13	13	5	Critically Low	No	Yes	Yes	No	No
17	1.500	10	13	6	Low	No	Yes	Yes	No	No
18	3.058	14	12	5	Critically Low	No	No	Yes	No	No
19	11.645	14	10	4	Critically Low	No	No	Yes	No	No
20	1.774	12	9	3	Critically Low	No	No	No	No	No
21	11.645	4	9	3	Critically Low	No	No	Yes	No	No
22	0.814	6	11	4	Critically Low	Yes	Yes	Yes	No	No
23	3.058	45	10	4	Critically Low	No	Yes	Yes	No	No
24	1.879	14	10	3	Critically Low	No	No	Yes	No	No
<b>Mean</b>	<b>3.249</b>	<b>13.7</b>	<b>10.4</b>	<b>4.1</b>	—	—	—	—	—	—
<b>Median</b>	<b>1.944</b>	<b>12.0</b>	<b>10.0</b>	<b>4.0</b>	—	—	—	—	—	—
<b>S.D.</b>	<b>3.337</b>	<b>10.2</b>	<b>1.7</b>	<b>1.0</b>	—	—	—	—	—	—
<b>Range</b>	<b>0.814 to 11.645</b>	<b>3 to 45</b>	<b>7 to 13</b>	<b>3 to 6</b>	—	—	—	—	—	—
<b>Frequency</b>	—	—	—	—	<b>21 Critically Low</b>	<b>12 Yes</b>	<b>8 Yes</b>	<b>21 Yes</b>	<b>0 Yes</b>	<b>0 Yes</b>
<b>Counts</b>	—	—	—	—	<b>3 Low</b>	<b>12 No</b>	<b>16 No</b>	<b>3 No</b>	<b>24 No</b>	<b>24 No</b>

JIF = Journal Impact Factor; PEDro = Physiotherapy Evidence Database (1 to 10 scale; higher value represents better study design).

AMSTAR 2 = A MeaSurement Tool to Assess systematic Reviews, version 2 (16 criteria; high value represents better review design).

SR = Systematic Review.

classified as having Spin being rated as no spin. No other trials were influenced by this methodological modification, as the remaining trials all had additional spin categories rated as positive. None of the SRs reported using a librarian to develop the literature search criteria for their SR or to perform the literature search. There were no findings of financial bias by the journal that published the SR. There was no identifiable conflict of interest for the RCTs contained within the SRs. There were, however, three instances where a conflict of interest was found for authors of an SR that was not disclosed.

### Relationships and differences of PEDro scores reported in SRs

The validity of the accuracy in reporting PEDro scores within each SR was assessed by comparing PEDro scores listed in the SRs versus the reported PEDro scores for the same articles on the PEDro website (Table 2). Overall, the SR PEDro scores and the website PEDro scores were well correlated using Pearson's  $r$  ( $r = 0.79$ ,  $p < .001$ ). The overall mean scores trended higher for the SRs ( $6.40 \pm 0.89$ ) compared to [https://](https://www.pedro.org.au/)

[www.pedro.org.au/](https://www.pedro.org.au/): ( $6.05 \pm 0.88$ ); however, the difference was not statistically significant ( $t = 1.354$ ,  $df = 46$ ,  $p = .18$ ). Of the 24 SRs, the <https://www.pedro.org.au/score> was higher 6 times and was lower 14 times, with four ties; this difference was statistically significant (Wilcoxon Sign Ranks test,  $p = .0153$ ).

### Journal impact factor, PEDro score, and AMSTAR 2

There were meaningful, statistically significant relationships between the median JIFs and mean PEDro scores ( $r = 0.55$ ,  $p = .0052$ ) and median JIFs and the percentage of low methodological quality clinical trials that were included in the 24 SRs ( $r = -0.73$ ,  $p < .001$ ) (Table 4). There were no statistically significant correlations for the variables of interest after alpha adjustment (Table 4).

## Discussion

### Quality of articles cited in SRs

Criterion 1 in the AMSTAR 2 requires that the research question for the SR is inclusive of the components of PICO (Population, Intervention, Comparator group,

**Table 4.** Preplanned correlations between variables related to the quality of the 24 systemic reviews (SR) and quality of articles that were reviewed by the SRs.

	Quality of SR	Quality of SR	Quality of SR	Quality of SR	Quality of Articles Reviewed by SR	Quality of Articles Reviewed by SR
	JIF (2019) of Journal that Published SR	AMSTAR 2: Number of 16 criteria met or partially met by SR	AMSTAR 2: Number of 7 Critical Validity Domain criteria met or partially met by SR	SR had 'Spin' (at least 2 of 7 criteria met)?	PEDro: Mean PEDro score (maximum = 10) of Articles Reviewed by SR (from PEDro website)	PEDro: % of Articles Reviewed by SR that were Low Quality (PEDro score < 6)
Quality of SR	<i>r</i> = 0.103 <i>p</i> = .631					
Quality of SR	<i>r</i> = 0.060 <i>p</i> = .782	--				
Quality of SR	<i>r</i> = -0.440 <i>p</i> = .0315	<i>r</i> = 0.000 <i>p</i> = .999	<i>r</i> = -0.041 <i>p</i> = .849			
Quality of Articles Reviewed by SR	<i>r</i> = 0.063 <i>p</i> = .768	<i>r</i> = -0.034 <i>p</i> = .872	<i>r</i> = -0.189 <i>p</i> = .376	<i>r</i> = 0.052 <i>p</i> = .808		
Quality of Articles Reviewed by SR	<i>r</i> = -0.087 <i>p</i> = .686	<i>r</i> = 0.005 <i>p</i> = .981	<i>r</i> = 0.261 <i>p</i> = .219	<i>r</i> = -0.059 <i>p</i> = .783	--	
Quality of Articles Reviewed by SR	<i>r</i> = 0.006 <i>p</i> = .978	<i>r</i> = 0.132 <i>p</i> = .539	<i>r</i> = -0.082 <i>p</i> = .702	<i>r</i> = -0.004778 <i>p</i> = .9823	<i>r</i> = 0.551 <i>p</i> = .0052	<i>r</i> = -0.726 <i>p</i> ≤ .001
Quality of Articles Reviewed by SR	<i>r</i> = 0.029 <i>p</i> = .892	<i>r</i> = -0.068 <i>p</i> = .753	<i>r</i> = 0.193 <i>p</i> = .367	<i>r</i> = -0.002 <i>p</i> = .992	<i>r</i> = -0.455 <i>p</i> = .0255	<i>r</i> = 0.456 <i>p</i> = .025
Quality of Articles Reviewed by SR	<i>r</i> = 0.348 <i>p</i> = .096	<i>r</i> = 0.061 <i>p</i> = .777	<i>r</i> = -0.079 <i>p</i> = .713	<i>r</i> = -0.275 <i>p</i> = .193	<i>r</i> = -0.398 <i>p</i> = .054	<i>r</i> = 0.127 <i>p</i> = .556
Quality of Articles Reviewed by SR	<i>r</i> = 0.095 <i>p</i> = .658	<i>r</i> = 0.173 <i>p</i> = .418	<i>r</i> = 0.102 <i>p</i> = .636	<i>r</i> = -0.092 <i>p</i> = .670	<i>r</i> = 0.077 <i>p</i> = .722	<i>r</i> = -0.004 <i>p</i> = .985
Quality of SR	<i>r</i> = -0.015 <i>p</i> = .946	<i>r</i> = 0.158 <i>p</i> = .461	<i>r</i> = 0.118 <i>p</i> = .582	<i>r</i> = -0.066 <i>p</i> = .759	<i>r</i> = -0.314 <i>p</i> = .136	<i>r</i> = -0.083 <i>p</i> = .700
Quality of SR	<i>r</i> = -0.2673 <i>p</i> = .2067	<i>r</i> = 0.307 <i>p</i> = .144	<i>r</i> = 0.175 <i>p</i> = .415	<i>r</i> = 0.177 <i>p</i> = .409	<i>r</i> = -0.488 <i>p</i> = .016	<i>r</i> = 0.412 <i>p</i> = .0452

Each column of correlations were considered to constitute a distinct family of analysis, and alpha was adjusted as per sequential method of Hochberg within each column. All *p* values less than 0.05 are shown in bold italic, but not all of these *p* values are considered statistically significant due to alpha adjustment. Only those values shown in bold, italicized, and underlined text are considered to be statistically significant after alpha adjustment.



and Outcome). Fourteen (58.3%) of the SRs included in this methodological review included clinical trials that did not meet criterion 1 on the PEDro scale regarding the inclusion and exclusion criteria of the RCT. Both criterion 1 in AMSTAR 2 and criterion 1 on the PEDro scale seek to identify who the study finding is generalizable to. If the 'population' of interest cannot be identified through reported inclusion and exclusion criteria of the primary research, it is not possible to determine if the identified RCT meets the PICO criteria for the SR. If the generalizability of the RCT cannot be specified, it should not be included in the SR.

Thirty-five percent of the clinical trials included in the SRs were low methodological quality. Seven (29%) of these SRs included pilot studies that were underpowered by their nature, and 7 (29%) included studies that were not peer-reviewed. One-third (33%) of the SRs based their conclusions at least in part to studies that were rated as low quality on the PEDro scale.

### **Quality of the SRs**

All SRs identified in this methodological SR represented 'Low' (3/24 SRs, 12.5%) to 'Critically Low' (21/24 SRs, 87.5%) ratings on the AMSTAR 2. Without diversity in quality on the AMSTAR 2 scoring across the SRs from critically low to high, it was not possible to draw conclusions regarding the factors that were related to these poor ratings. Our findings are similar to the results of Almeida et al., who found that 16% and 74% of the SRs were rated Low and Critically Low, respectively, in an appraisal of systematic reviews on exercise therapy for chronic low back pain [33].

An important and critical consideration for the AMSTAR 2 is the prospective registration of the SR. Fifteen (63%) of the SRs included in this study did not fulfill this criterion. Given that reviewers are blinded during the review process, it becomes essential that journal editors verify that an SR meets this criterion prior to sending an SR out for review.

AMSTAR 2 criterion 4 is a critical metric and asks the question, 'Did the review authors use a comprehensive literature search strategy?' It is widely accepted that the inclusion of a professional librarian with expertise in SR search methodology decreases errors in search strategies and improves the quality of SRs [22–25,36,37]. Beyond these factors, it may be essential to include a professional librarian that has no interest in the outcome of the SR to decrease the risk of bias when creating search strategies. Although the use of a professional librarian is a gold standard methodological practice when performing an SR with or without a meta-analysis, none of the SRs in this methodological review reported the inclusion of a professional librarian.

AMSTAR 2 criterion 13, a critical domain, specifies that authors account for the risk of bias (RoB) in

individual studies when interpreting/discussing SR results. More than half of the SRs included in this methodological review (14/24, 58%) did not meet this criterion. It has been suggested that either the PEDro score or the CRoB tool can be used for this purpose [38]. While our study only considered studies using the PEDro score in isolation, we would argue that in the absence of another tool, the best way to account for a high risk of bias during the SR process would be to exclude clinical trials that have low methodological quality (PEDro score <6). Given that the updated AMSTAR 2 recommendations were published in 2017 [17], it may not be fair to consider the SRs published before that time that was included in this methodological review through that lens. It should, however, be an important consideration moving forward.

### **Conflict of interest, journal financial bias, and spin**

Although none of SRs were determined to be published in financially biased journals, 3 SRs (12.5%) had authors with undisclosed conflicts of interest, and 12 (50%) SRs reported conclusions that were inconsistent with the SR results ('spin'). Our findings regarding 'spin' were in contrast to the findings of Nascimento et al., who found that 80% of the abstracts in SRs for physical therapy interventions for low back pain contained 'spin' [39]. In the present study, no associations were found between 'spin' and SR quality or the quality of the articles reviewed by the SRs. Editors and reviewers should be diligent in ensuring that the conclusions of SRs are factually consistent with the results.

### **Journal impact factor, PEDro score, and AMSTAR 2**

There were moderate to strong relationships between the median JIFs, mean PEDro scores, and percent of PEDro scores representing low methodological quality for the clinical trials included in the SRs. This suggests that there may be a relationship between clinical trial methodological quality and JIF. This finding was surprising, given that a recent systematic review found that there was a weak and inconsistent relationship between JIF and methodological quality in randomized clinical trials [40]. A potential reason that our findings are different from previous literature is that we chose to use median JIFs for our statistical analyses. The use of the mean as a measure of central tendency for JIFs may be problematic secondary to a small number of JIFs that may be extraordinarily high that cause an over-inflation of the mean as a measure of central tendency. Nascimento et al. found that JIFs for the journals that published the SRs were not associated with the methodological quality of the SRs reported using the AMSTAR 2 [41]. Our findings support this conclusion. This finding is also surprising. In 2005 it was suggested that, on a per study basis, systematic

reviews are cited more frequently than the research that was used to create them [42]. This finding has been echoed by publishers who recommend that journals publish more review articles as a strategy for increasing the journals' JIF [43]. Our study included SRs that have been published in the last 5 years for musculoskeletal physical therapy interventions. The level of evidence may no longer be as important as the quality of evidence for attaining JIFs in physical therapy journals.

### **Accuracy of PEDro scores reported in SRs**

There was a meaningful, statistically significant correlation between PEDro score attained by authors of the SRs and the PEDro scores achieved from the PEDro website. Although the scores differed slightly, there were no statistically significant differences found. This suggests that authors not explicitly trained in the PEDro scoring methodology are able to attain similar scores without introducing bias.

### **Summary of evidence**

Slightly over 58% of the SRs included RCTs that did not report inclusion and exclusion criteria (criterion 1 on the PEDro), and 35% percent of articles cited by the SRs were low quality on the PEDro scale. Twenty-one of 24 SRs (87.5%) included in this study were rated as critically low. The majority (63%) of the SRs were not prospectively registered, 50% reported conclusions that were inconsistent with their results, and none utilized a professional librarian.

### **Limitations**

The major limitation of our study is the generalizability of our findings. Our review was interested in controlling for the wide variability of single instruments and the combination of those instruments that are used to critically appraise RCTs during the SR process. Our findings are only generalizable to SRs that used the PEDro score as a single instrument for the critical appraisal of RCTs during the SR process. It is possible that by only selecting articles that used the PEDro scale, our SR may have excluded higher quality SRs that may have used the CROB tool. Future research could assess SRs using other quality criteria such as the CROB tool. To attain a reasonable number of research articles to allow for a practical literature review and statistical analysis, we limited our literature search to SRs that included physical therapy musculoskeletal interventions by publication date. Our findings, therefore, cannot be generalized to SRs outside our search criteria.

## **Conclusions**

All studies included in this methodological SR included SRs that represented Level 1A evidence related to a Grade A practice recommendation [1], considered to be strong recommendations that clinicians should follow unless they have a 'clear and compelling' reason not to follow them [1]. Based on the findings of this methodological SR, such recommendations may be problematic. The results of this methodological SR suggest that the highest level of evidence may over-represent poor quality evidence based on how the SRs and the RCTs contained within those SRs were reported. There were 19 different combinations of critical appraisal tools used to critically appraise RCTs included within the identified SRs based on our preliminary search. It may be time to standardize how RCTs are critically assessed during the SR process and reconsider how level and quality of evidence should be used when integrating research into clinical practice.

### **Disclosure statement**

No potential conflict of interest was reported by the authors.

### **Funding**

No funding was provided for the development of this manuscript. I certify that neither I nor the co-investigators on this research project have any conflicts of interest related to this work.

### **Notes on contributors**

**Sean P. Riley**, PT, DPT, ScD is an Assistant Professor in the Physical Therapy program at Sacred Heart University in Fairfield, CT, USA and faculty in the Sacred Heart University Orthopedic Physical Therapy Residency program. He graduated from the UConn Physical Therapy Program in 2004 and completed his Doctorate in Physical Therapy at Simmons College in 2006. He completed his Doctor of Science from Texas Tech University Health Science Center in 2013. Sean became board certified in Orthopedics through the American Board of Physical Therapy Specialties in 2007 and completed his sports certification in 2009. He completed his Fellowship in Manual Therapy from Regis University in 2018.

**Brian T. Swanson**, PT, DSc is an Assistant Professor of Physical Therapy in the Department of Rehabilitation Sciences at The University of Hartford in West Hartford, CT, USA and faculty in the Select Medical Orthopedic Physical Therapy Residency program. Dr. Swanson graduated from Quinnipiac University with a Bachelor of Science degree in Physical Therapy in 1996 and earned a Doctor of Science in Physical Therapy from Andrews University in 2012. Brian became board certified in Orthopedics through the American Board of Physical Therapy Specialties in 2005. He completed a two-year residency and Fellowship in the Kaltenborn-Evjenth Norwegian Approach to Orthopaedic Manual Physical Therapy at the Institute of Orthopaedic Manual Therapy, Woburn, MA in 2006.

**Steven F. Sawyer**, PT, PhD, is professor and chair of the Department of Rehabilitation Sciences, Associate Dean for Faculty Development, and director of the Clinical Postural Control Research Laboratory in the Center for Rehabilitation Research, all within the School of Health Professions at Texas Tech University Health Sciences Center. His PhD in Neuroscience was earned from the University of California at San Diego in 1989. He graduated from PT school at the Texas Tech University Health Sciences Center in 1997 and has been on the faculty since then. He received certification in Vestibular Rehabilitation from Emory University in 2001.

**Jean-Michel Brismée**, PT, ScD is a professor in the Doctor of Science Program in Physical Therapy at Texas Tech University Health Sciences Center in Lubbock, Texas. Dr. Brismée graduated from the Catholic University of Louvain-la-Neuve in Belgium with Bachelor of Science degrees in Physical Education and Physical Therapy in 1985. He graduated from Texas Tech University in 1996 with a Master of Sciences in Sports Health and earned a Doctor of Science degree in Physical Therapy in 2003. He is the Chair of the Research Committee and Director of the Fellowship program in Orthopaedic Manual Therapy of the IAOM-US, Editor-in-Chief of the *Journal of Manual and Manipulative Therapy* and maintains clinical practical at University Medical Center in Lubbock, Texas.

**Geoffrey Staysniak** is Health Sciences Librarian at Sacred Heart University. He has a BA from Fairfield University, an MLIS from Simmons College.

## ORCID

Sean Riley  <http://orcid.org/0000-0002-8854-8024>

Jean-Michel Brismée  <http://orcid.org/0000-0002-1037-7704>

Geoffrey Staysniak  <http://orcid.org/0000-0002-8121-9164>

## References

- [1] Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011 Jul;128(1):305–310. PubMed PMID: 21701348; PubMed Central PMCID: PMC3124652.
- [2] Zadro J, O'Keeffe M, Maher C. Do physical therapists follow evidence-based guidelines when managing musculoskeletal conditions? Systematic review. *BMJ Open*. 2019 Oct 7;9(10):e032329. PubMed PMID: 31591090; PubMed Central PMCID: PMC6797428.
- [3] Mandel D, Littner Y, Mimouni FB, et al. Conclusiveness of the Cochrane neonatal reviews: a systematic analysis. *Acta Paediatr*. 2006 Oct;95(10):1209–1212. PubMed PMID: 16982491.
- [4] Cohen S, Lubetzky R, Mimouni FB, et al. Conclusiveness of the Cochrane reviews in pediatric-gastroenterology: a systematic analysis. *Eur J Gastroenterol Hepatol*. 2013 Feb;25(2):252–254. PubMed PMID: 23044810.
- [5] Cohen S, Mandel D, Mimouni FB, et al. Conclusiveness of the Cochrane reviews in nutrition: a systematic analysis. *Eur J Clin Nutr*. 2014 Feb;68(2):143–145. PubMed PMID: 24327125.
- [6] Mimouni M, Mimouni F, Segev F. Conclusiveness of the Cochrane eye and vision group reviews. *BMC Res Notes*. 2015 June 16;8:242. PubMed PMID: 26076817; PubMed Central PMCID: PMC4468979.

- [7] Yin S, Chuai Y, Wang A, et al. Conclusiveness of the Cochrane reviews in gynaecological cancer: A systematic analysis. *J Int Med Res*. 2015 Jun;43(3):311–315. PubMed PMID: 25870179.
- [8] Zhang X, Wu Z, Zhao H, et al. Conclusiveness of the Cochrane reviews in palliative and supportive care for cancer. *Am J Hosp Palliat Care*. 2017 Feb;34(1):53–56. PubMed PMID: 26353985.
- [9] Momosaki R, Tsuboi M, Yasufuku Y, et al. Conclusiveness of Cochrane Reviews in physiotherapy: a systematic search and analytical review. *Int J Rehabil Res*. 2019 Jan 16. DOI:10.1097/MRR.0000000000000338. PubMed PMID: 30664013.
- [10] Uttley L, Montgomery P. The influence of the team in conducting a systematic review. *Syst Rev*. 2017 Aug 1;6(1):149. PubMed PMID: 28764779; PubMed Central PMCID: PMC5540536.
- [11] Oliveira CB, Elkins MR, Lemes IR, et al. A low proportion of systematic reviews in physical therapy are registered: a survey of 150 published systematic reviews. *Braz J Phys Ther*. 2018 May - Jun;22(3):177–183. PubMed PMID: 29128407; PubMed Central PMCID: PMC5993937.
- [12] Riley SP, Swanson B, Brismee JM, et al. A systematic review of orthopaedic manual therapy randomized clinical trials quality. *J Man Manip Ther*. 2016 Dec;24(5):241–252. PubMed PMID: 27956817; PubMed Central PMCID: PMC5125432.
- [13] Gentles SJ, Charles C, Nicholas DB, et al. Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research. *Syst Rev*. 2016 Oct 11;5(1):172. PubMed PMID: 27729071; PubMed Central PMCID: PMC5059917.
- [14] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009 Jul;21(339):b2700. PubMed PMID: 19622552; PubMed Central PMCID: PMC2714672.
- [15] Green S, Higgins JPT, Alderson P, et al. Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011 [Cited 2019 Oct 2]. Available from: <https://training.cochrane.org/handbook>
- [16] Shamseer L, Hopewell S, Altman DG, et al. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials*. 2016 June 24;17(1):301. PubMed PMID: 27343072; PubMed Central PMCID: PMC4921029.
- [17] Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017 Sep 21;358:j4008. PubMed PMID: 28935701; PubMed Central PMCID: PMC5833365 at [http://www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.
- [18] Nascimento DP, Gonzalez GZ, Araujo AC, et al. Eight out of every ten abstracts of low back pain systematic reviews presented spin and inconsistencies with the full text: an analysis of 66 systematic reviews. *J Orthop*

- Sports Phys Ther. 2019 August;23(1):34. PubMed PMID: 31443622.
- [19] Furland AD, Malmivaara A, Chou R, et al. Updated method guideline for systematic reviews in the Cochrane back and neck group. *Spine (Phila Pa 1976)*. 2015;40(21):1660–1673.
  - [20] Lefebvre C, Manheimer E, Glanville J. Searching for studies. *Cochrane handbook for systematic reviews of interventions*. Chichester (UK): John Wiley & Sons, Ltd; 2008. p. 95–150.
  - [21] McGowan J, Sampson M, Salzwedel DM, et al. Press peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75(Supplement C): 40–46.
  - [22] Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, et al. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015 Jun;68 (6):617–626. PubMed PMID: 25766056.
  - [23] Meert D, Torabi N, Costella J. Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. *J Med Libr Assoc*. 2016 Oct;104(4):267–277. PubMed PMID: 27822147; PubMed Central PMCID: PMC665079487.
  - [24] Koffel JB. Use of recommended search strategies in systematic reviews and the impact of librarian involvement: a cross-sectional survey of recent authors. *PLoS One*. 2015;10(5):e0125931. PubMed PMID: 25938454; PubMed Central PMCID: PMC665079487.
  - [25] Li L, Tian J, Tian H, et al. Network meta-analyses could be improved by searching more sources and by involving a librarian. *J Clin Epidemiol*. 2014 Sep;67 (9):1001–1007. PubMed PMID: 24841794.
  - [26] Maher CG, Sherrington C, Herbert RD, et al. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003 Aug;83(8):713–721. PubMed PMID: 12882612.
  - [27] Foley NC, Bhogal SK, Teasell RW, et al. Estimates of quality and reliability with the physiotherapy evidence-based database scale to assess the methodology of randomized controlled trials of pharmacological and nonpharmacological interventions. *Phys Ther*. 2006 Jun;86(6):817–824. PubMed PMID: 16737407.
  - [28] Macedo LG, Elkins MR, Maher CG, et al. There was evidence of convergent and construct validity of Physiotherapy Evidence Database quality scale for physiotherapy trials. *J Clin Epidemiol*. 2010 Aug;63 (8):920–925. PubMed PMID: 20171839.
  - [29] Sherrington C, Herbert RD, Maher CG, et al. PEDro. A database of randomized trials and systematic reviews in physiotherapy. *Man Ther*. 2000 Nov 5; (4):223–226. DOI:10.1054/math.2000.0372. PubMed PMID: 11052901.
  - [30] Yamato TP, Maher C, Koes B, et al. The PEDro scale had acceptably high convergent validity, construct validity, and interrater reliability in evaluating methodological quality of pharmaceutical trials. *J Clin Epidemiol*. 2017 Jun;86:176–181. PubMed PMID: 28288916.
  - [31] de Almeida MO, Saragiotto BT, Maher C, et al. Allocation concealment and intention-to-treat analysis do not influence the treatment effects of physical therapy interventions in low back pain trials: a meta-epidemiologic study. *Arch Phys Med Rehabil*. 2019 Jul;100(7):1359–1366. PubMed PMID: 30710510.
  - [32] Riley SP, Swanson BT, Brismee JM, et al. Low reproducibility of randomized clinical trials methodology related to sampling: a systematic methodological review. *J Man Manip Ther*. 2019 Dec;27(5):258–266. PubMed PMID: 30935323; PubMed Central PMCID: PMC6650791.
  - [33] Almeida MO, Yamato TP, Parreira P, et al. Overall confidence in the results of systematic reviews on exercise therapy for chronic low back pain: a cross-sectional analysis using the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) 2 tool. *Braz J Phys Ther*. 2020 Mar - Apr;24 (2):103–117. PubMed PMID: 31113734; PubMed Central PMCID: PMC6650791.
  - [34] Think Check Submit. 2020 [Cited 2020 Jan 22]. Available from: <https://thinkchecksubmit.org/check/>
  - [35] Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75:800–802.
  - [36] Salvador-Olivan JA, Marco-Cuenca G, Arquero-Aviles R. Errors in search strategies used in systematic reviews and their effects on information retrieval. *J Med Libr Assoc*. 2019 Apr;107(2):210–221. PubMed PMID: 31019390; PubMed Central PMCID: PMC6650791.
  - [37] Opheim E, Andersen PN, Jakobsen M, et al. Poor quality in systematic reviews on PTSD and EMDR - An examination of search methodology and reporting. *Front Psychol*. 2019;10:1558. PubMed PMID: 31354575; PubMed Central PMCID: PMC6650791.
  - [38] Moseley AM, Rahman P, Wells GA, et al. Agreement between the Cochrane risk of bias tool and Physiotherapy Evidence Database (PEDro) scale: A meta-epidemiological study of randomized controlled trials of physical therapy interventions. *PLoS One*. 2019;14(9):e0222770. PubMed PMID: 31536575; PubMed Central PMCID: PMC6650791.
  - [39] Nascimento DP, Gonzalez GZ, Araujo AC, et al. Eight out of every ten abstracts of low back pain systematic reviews presented spin and inconsistencies with the full text: an analysis of 66 systematic reviews. *J Orthop Sports Phys Ther*. 2020 Jan;50(1):17–23. PubMed PMID: 31443622.
  - [40] Saginur M, Fergusson D, Zhang T, et al. Journal impact factor, trial effect size, and methodological quality appear scantily related: a systematic review and meta-analysis. *Syst Rev*. 2020 March 9;9(1):53. PubMed PMID: 32164791; PubMed Central PMCID: PMC6650791.
  - [41] Nascimento DP, Gonzalez GZ, Araujo AC, et al. Journal impact factor is associated with PRISMA endorsement, but not with the methodological quality of low back pain systematic reviews: a methodological review. *Eur Spine J*. 2020 Mar;29(3):462–479. PubMed PMID: 31707453.
  - [42] Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA*. 2005 May 18;293(19):2362–2366. PubMed PMID: 15900006.
  - [43] Increasing citations and improving your impact factor. Sage Publishing; 2020 [Cited 2020 May 26]. Available from: <https://us.sagepub.com/en-us/nam/increasing-citations-and-improving-your-impact-factor>



## Appendix 1: Electronic Search Strategies

### PubMed

(musculoskeletal system[mh:noexp] OR tendons[mh:noexp] OR rotator cuff[mh] OR patellar ligament[mh] OR hamstring tendons[mh] OR Achilles tendon[mh] OR skeleton[mh:noexp] OR joints[mh:noexp] OR zygapophyseal joint[mh] OR temporomandibular joint[mh] OR sternocostal joints[mh] OR sternoclavicular joint[mh] OR shoulder joint[mh] OR sacroiliac joint[mh] OR pubic symphysis[mh] OR ligaments, articular[mh:noexp] OR round ligament of femur[mh] OR palmar plate[mh] OR posterior cruciate ligament[mh] OR plantar plate[mh] OR longitudinal ligaments[mh] OR ligamentum flavum[mh] OR collateral ligaments[mh:noexp] medial collateral ligament, knee[mh] OR lateral ligament, ankle[mh] OR collateral ligament, ulnar[mh] OR anterior cruciate ligament[mh] OR knee joint[mh:noexp] OR patellofemoral joint[mh] OR menisci, tibial[mh] OR joint capsule[mh:noexp] OR synovial membrane[mh:noexp] OR hip joint[mh:noexp] OR hand joints[mh:noexp] OR wrist joint[mh:noexp] OR triangular fibrocartilage[mh] OR metacarpophalangeal joint[mh:noexp] OR finger joint[mh] OR carpometacarpal joints[mh] OR carpal joints[mh] OR foot joints[mh:noexp] OR toe joint[mh:noexp] OR tarsal joints[mh:noexp] OR subtalar joint[mh] OR metatarsophalangeal joint[mh:noexp] OR ankle joint[mh] OR elbow joint[mh] OR bursa, synovial[mh] OR atlanto-occipital joint[mh] OR atlanto-axial joint[mh] OR acromioclavicular joint OR bone and bones[mh:noexp] OR spine[mh:noexp] [mh] OR thoracic vertebrae[mh] OR spinal canal[mh] OR sacrum[mh] OR lumbar vertebrae[mh] OR coccyx[mh] OR cervical vertebrae[mh:noexp] cervical atlas[mh] OR axis, cervical vertebra[mh:noexp] OR sesamoid bones[mh:noexp] OR patella[mh] OR rib cage[mh:noexp] OR sternum[mh:noexp] OR xiphoid bone[mh] OR manubrium[mh] OR ribs[mh:noexp] cervical rib[mh] OR hyoid bone[mh] OR epiphyses[mh:noexp] OR growth plate[mh] OR diaphyses[mh] OR bones of upper extremity[mh:noexp] OR scapula[mh:noexp] OR glenoid cavity[mh] OR coracoid process[mh] OR acromion[mh] OR hand bones[mh:noexp] OR clavicle[mh] OR arm bones[mh:noexp] OR ulna[mh] OR radius[mh] OR humerus[mh:noexp] OR humeral head[mh] OR bones of lower extremity[mh:noexp] OR muscles[mh:

noexp] OR muscle, smooth[mh:noexp] OR muscle, smooth, vascular [mh] OR muscle, skeletal[mh:noexp] OR respiratory muscles[mh:noexp] OR intercostal muscles[mh] OR diaphragm[mh] OR quadriceps muscle[mh] OR psoas muscles[mh] OR pharyngeal muscles[mh] OR pectoralis muscles[mh] OR neck muscles[mh] OR masticatory muscles [mh:noexp] OR temporal muscle[mh] OR pterygoid muscles[mh] OR masseter muscle[mh] OR hamstring muscles[mh] OR gracilis muscle[mh] OR facial muscles[mh] OR deltoid muscle[mh] OR back muscles[mh:noexp] OR superficial back muscles[mh] OR paraspinal muscles[mh] OR intermediate back muscles[mh] OR abdominal muscles[mh:noexp] OR rectus abdominis[mh] OR pelvic floor[mh] OR abdominal oblique muscles[mh] OR ligaments[mh:noexp] OR round ligaments[mh] ligaments, articular[mh:noexp] OR posterior cruciate ligament[mh] OR broad ligament[mh] OR fascia [mh:noexp] OR tenon capsule[mh] OR fascia lata[mh] OR cartilage[mh:noexp] OR hyaline cartilage[mh] OR fibrocartilage[mh] OR fibrocartilage[mh:noexp] OR meniscus [mh:noexp] OR menisci, tibial OR intervertebral disc [mh] OR elastic cartilage[mh:noexp] OR costal cartilage[mh] OR cartilage, articular[mh] AND (physical therap\* OR physiotherap\* OR physical therapy modalities[mh:noexp]) OR musculoskeletal manipulations[mh:noexp] OR therapy, soft tissue[mh:noexp] OR motion therapy, continuous passive OR manipulation, spinal[mh] manipulation, orthopedic[mh] OR manipulation, osteopathic[mh] OR extracorporeal shockwave therapy[mh] OR exercise therapy[mh: noexp] OR muscle-stretching exercises[mh] OR exercise movement techniques [mh:noexp] OR breathing exercise[mh] OR electric stimulation therapy[mh:noexp] OR transcutaneous electric nerve stimulation[mh] OR spinal cord stimulation[mh] OR pulsed radiofrequency treatment[mh] OR electroacupuncture[mh] OR physical conditioning, human[mh:noexp] OR resistance training [mh] OR plyometric exercise[mh] OR high-intensity interval training[mh] OR circuit-based exercise[mh] OR exercise[mh:noexp] OR warm-up exercise[mh] OR walking[mh: noexp] OR swimming[mh] OR cryotherapy[mh] OR traction [mh] OR running[mh:noexp] OR early ambulation[mh] OR pain science education OR pain neuroscience education OR neuroscience education) AND (systematic review\*) AND ('2016/01/01'[PDat]: '2019/12/12'[PDat]) AND (English[lang])